

Understanding One-Way ANOVA: A Step-by-Step Guide Using Stata

Authored by
Mohammed looti

November 8, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Understanding One-Way ANOVA: A Step-by-Step Guide Using Stata*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13768>

Introduction to the One-Way Analysis of Variance (ANOVA)

A [One-Way ANOVA](#) (Analysis of Variance) is a powerful inferential statistical test utilized to assess whether there is a statistically significant difference between the means of three or more independent, unrelated groups. This technique is foundational in fields ranging from psychology and medicine to economics, providing a structured approach to comparing group variations. The core principle of the [ANOVA](#) is the partitioning of the total variability observed in the data into two components: variability explained by the differences between the group means (treatment effect) and unexplained variability within the groups (error).

The designation of "one-way" signifies that the analysis involves only one categorical independent variable, often referred to as a factor. This factor must have at least three levels or groups. If the study were instead interested in the simultaneous effects of two or more independent factors on a single response variable, a [Two-Way ANOVA](#) or a more complex factorial design would be necessary. Understanding the distinction between these variants is crucial for selecting the appropriate statistical tool for hypothesis testing.

The primary objective of conducting a [One-Way ANOVA](#) is to test the null hypothesis (**H₀**), which posits that the population means of all groups are equal. The alternative hypothesis (**H_a**) is that at least one group mean is different from the others. Before proceeding with the analysis in [Stata](#), it is essential to ensure that the data meets key assumptions, including the normality of residuals, the independence of observations, and, critically, the homogeneity of variances (Levene's test). This tutorial will guide you through the process of executing and interpreting this test using the statistical software package [Stata](#).

Setting Up the Analysis in Stata: Loading and Exploring Data

For this practical illustration, we will utilize a built-in Stata dataset specifically designed for educational purposes, named **systolic**. This dataset is ideally suited for demonstrating the application of the [One-Way ANOVA](#), as it contains measurements related to a clinical intervention involving multiple treatment groups. The dataset encompasses observations from 58 distinct individuals and contains several variables pertinent to our analysis.

The variables central to our investigation within the **systolic** dataset are structured as follows:

Drug used: This is our categorical independent variable (the factor), representing the four different drug treatments administered.

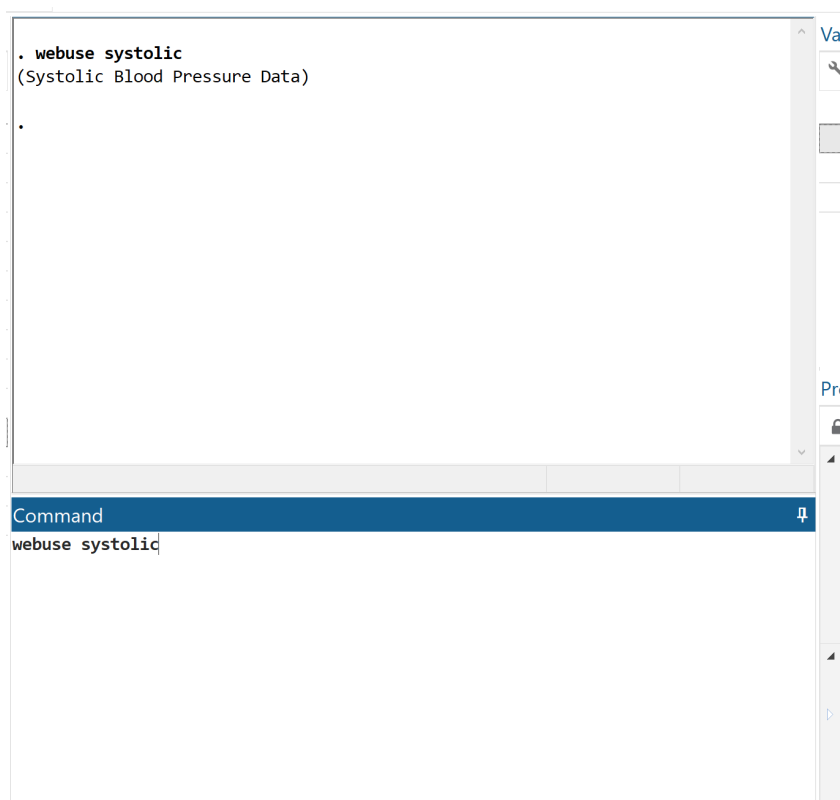
Patient's disease: An additional categorical variable that might be relevant for other analyses but is not the factor of interest in this specific one-way test.

Change in systolic blood pressure: This is the continuous dependent or response variable, which is the outcome we are measuring across the different drug groups.

Our goal is to determine if the variation in the **Drug used** factor leads to a statistically significant impact on the **Change in systolic blood pressure**. The initial procedural steps involve accessing and reviewing the data within the Stata environment to ensure familiarity with the data structure.

Step 1: Load the data.

The first action required in Stata is to load the sample dataset. This is executed by typing the command **webuse systolic** directly into the command line interface and pressing Enter. This command retrieves the dataset from the Stata server, making it available in the current session's memory for subsequent analysis.



Step 2: View the raw data.

Before launching into formal statistical tests, it is considered best practice to examine the raw data structure. This visual inspection helps confirm variable types, identify potential data entry errors, or note any unusual patterns. To view the data editor in browse mode, navigate through the top menu bar by selecting **Data > Data Editor > Data Editor (Browse)**. This action will display the actual data matrix containing the observations for all 58 patients.

Data Editor (Browse) - [systolic]

File Edit View Data Tools

systolic[1] 42

	drug	disease	systolic				
1	1	1	42				
2	1	1	44				
3	1	1	36				
4	1	1	13				
5	1	1	19				
6	1	1	22				
7	1	2	33				
8	1	2	26				
9	1	2	33				
10	1	2	21				
11	1	3	31				
12	1	3	-3				
13	1	3	25				
14	1	3	25				
15	1	3	24				
16	2	1	28				
17	2	1	23				
18	2	1	34				

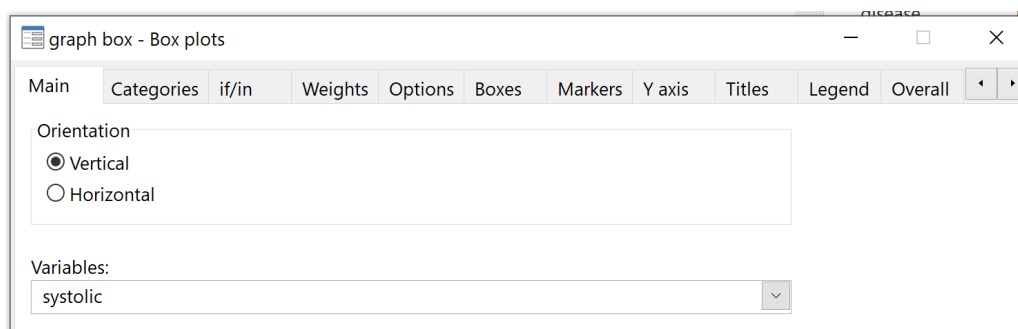
Visualizing Group Distributions Using Box Plots

Data visualization is an indispensable preliminary step in any statistical analysis. Generating graphical summaries allows researchers to gain intuitive insight into the central tendency, spread, and symmetry of the data within each group before the formal hypothesis test is applied. For comparing the distributions of a continuous variable across multiple categorical groups, the [Box plot](#), also known as a box-and-whisker plot, is an exceptionally effective tool.

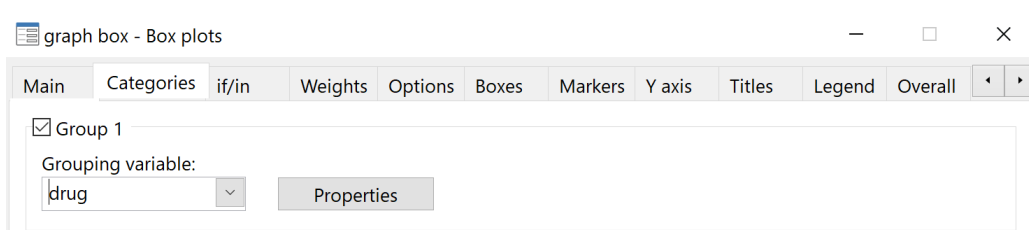
Step 3: Visualize the data.

We will generate multiple [Box plot](#) graphs to view the distribution of systolic blood pressure change values for each distinct category of the administered drug. This visualization step helps us assess, visually, whether the means appear different and if the variance (spread) seems consistent across the drug groups--a crucial check related to the ANOVA assumption of homogeneity of variances.

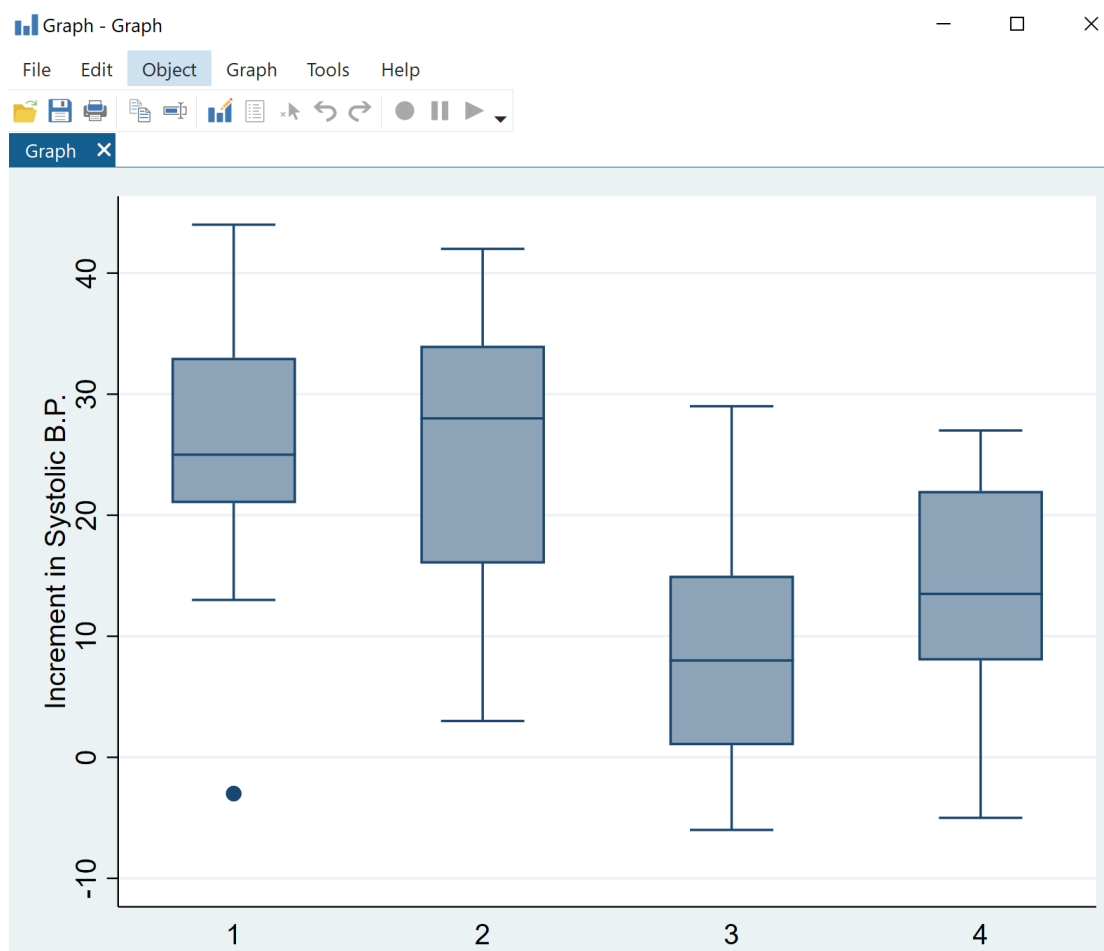
To create these plots in Stata using the graphical user interface (GUI), follow these steps: navigate to the top menu bar, select **Graphics > Box plot**. When prompted, select **Systolic** as the variable to be plotted. You must also specify that the box plots should be generated separately for each level of the drug variable, allowing for a side-by-side comparison of the distributions.



Further configuration, often involving specifying the categorical variable that defines the groups (in this case, **drug**), is necessary to ensure the plots are correctly grouped. After defining the response variable (Systolic) and the grouping variable (Drug), click **OK**.



Upon execution, a graphic containing four separate boxplots will be displayed, one corresponding to each drug category. Observing this visualization immediately suggests that the center (median, indicated by the line inside the box) and the overall distribution of changes in systolic blood pressure appear to vary substantially among the drug categories. While the visual evidence hints at mean differences, the [ANOVA](#) statistical test is required to formally determine if these observed differences are large enough to be considered statistically significant, rather than merely due to random sampling fluctuation.



Executing the One-Way ANOVA and Interpreting Primary Results

With the data loaded and visually inspected, we proceed to the formal hypothesis test using the [One-Way ANOVA](#) procedure in Stata. This procedure calculates the **F-ratio**, which is the ratio of variance explained by the factor (between-group variance) to the variance unexplained (within-group variance). A larger F-ratio suggests that the differences between the group means are substantial relative to the inherent variability within the groups.

Step 4: Perform a one-way ANOVA.

To initiate the ANOVA analysis through the Stata GUI, navigate to the following menu path: **Statistics > Linear models and related > ANOVA/MANOVA > One-Way ANOVA**. In the resulting dialog box, specify the variables: the **response variable** should be set to **systolic** (the outcome measure), and the **factor variable** (the independent group variable) should be set to **drug**. Furthermore, it is highly recommended to check the box next to **Produce summary table**. This ensures that the output includes descriptive statistics (means, standard deviations, and group sizes) for each drug category, which aids in result interpretation and reporting. Finally, click **OK** to

run the test.

The screenshot shows the 'oneway - One-way analysis of variance' dialog box in Stata. The 'Main' tab is selected. The 'Response variable:' is set to 'systolic' and the 'Factor variable:' is set to 'drug'. Under 'Multiple-comparison tests', the options 'Bonferroni', 'Scheffe', and 'Sidak' are all unchecked. In the 'Output' section, 'Produce summary table' is checked, while 'Suppress means', 'Suppress standard deviations', 'Suppress frequencies', and 'Suppress number of obs.' are unchecked. In the 'Other' section, 'Suppress the ANOVA table', 'Show numeric codes, not labels', and 'Do not break wide tables' are all unchecked. At the bottom, the 'Treat missing values as categories' checkbox is unchecked. The 'OK' button is highlighted with a blue border.

The resulting output table presents the core statistics of the [ANOVA](#) test, including the sources of variation (Between groups and Within groups), the sum of squares, degrees of freedom, mean squares, and the final inferential statistics. The critical values for interpreting the test outcome are the **F-statistic** and its corresponding **p-value**.

. oneway systolic drug, tabulate

Drug Used	Summary of Increment in Systolic B.P.				
	Mean	Std. Dev.	Freq.		
1	26.066667	11.677002	15		
2	25.533333	11.61813	15		
3	8.75	10.0193	12		
4	13.5	9.3238047	16		
Total	18.87931	12.800874	58		

Source	Analysis of Variance				
	SS	df	MS	F	Prob > F
Between groups	3133.23851	3	1044.41284	9.09	0.0001
Within groups	6206.91667	54	114.942901		
Total	9340.15517	57	163.862371		

In this example, the calculated [F-statistic](#) is **9.09**, associated with a [p-value](#) of **0.0001**. When evaluating the results, we compare the [p-value](#) to a predetermined significance level, typically **alpha = 0.05**. Since our observed [p-value](#) (0.0001) is significantly smaller than 0.05, we reject the null hypothesis. The rejection of **H₀** leads to the conclusion that there is sufficient statistical evidence to state that the mean change in systolic blood pressure is not equal across all drug groups. Crucially, however, the [One-Way ANOVA](#) only tells us that *at least two* of the group means differ; it does not specify which particular pairs are significantly different. For this detailed information, we must proceed to a post-hoc analysis.

Post-Hoc Analysis: Identifying Specific Group Differences (Tukey's HSD)

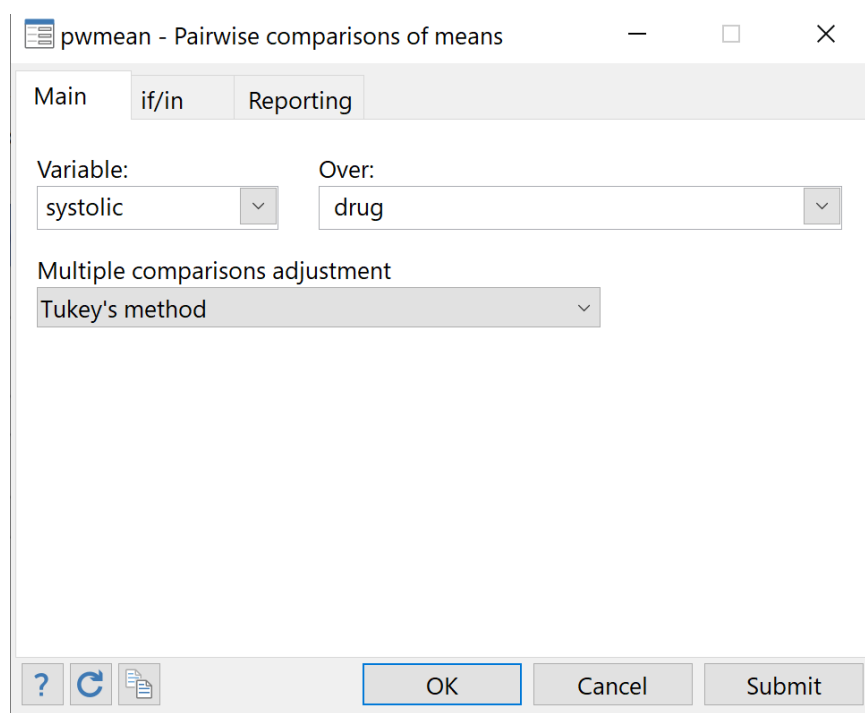
Once the overall ANOVA test indicates a significant difference among the group means, the next logical step is to perform post-hoc tests, or multiple comparison tests. These tests are necessary to pinpoint the exact pairs of groups that differ significantly from each other, allowing for detailed conclusions regarding the specific effectiveness of the drug treatments. Performing these pairwise comparisons without adjustment increases the risk of Type I error (falsely rejecting the null hypothesis), making an adjustment method essential.

Step 5: Perform multiple comparison tests.

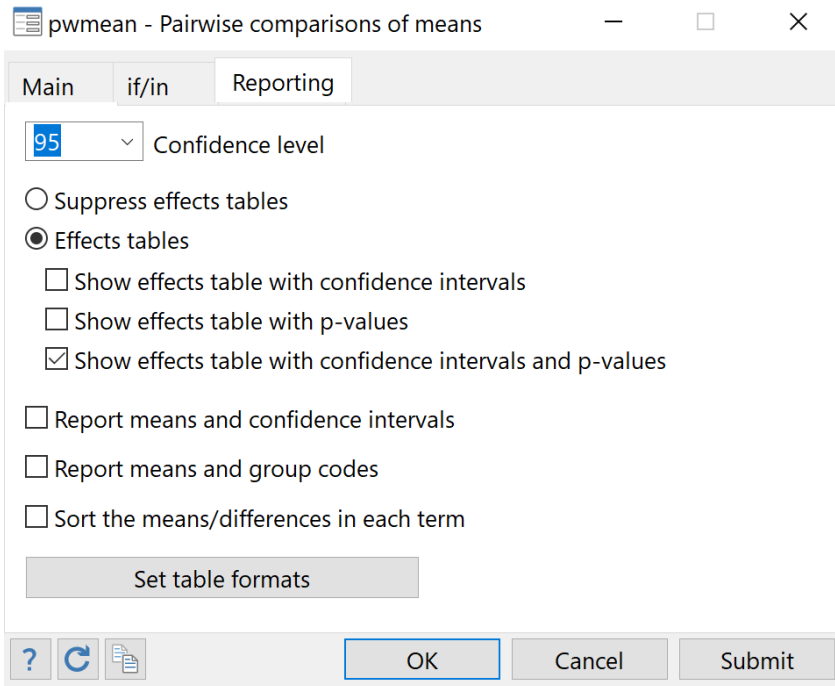
We will employ [Tukey's method](#) (also known as Tukey's Honestly Significant Difference or HSD test). This method is widely used because it controls the family-wise error rate, ensuring that the

overall probability of making at least one Type I error across all possible pairwise comparisons remains at the desired alpha level (e.g., 0.05).

In Stata, access the post-hoc tests by navigating to: **Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Pairwise comparisons of means**. In the dialog box, set the **Variable** (response variable) to **systolic** and the **Over** variable (explanatory factor) to **drug**. Under the **Multiple comparisons adjustment** option, select **Tukey's method**.



To ensure the output is detailed and suitable for academic reporting, navigate to the **Reporting** subheading within the dialog box. Click the button next to **Effects tables** and ensure the box labeled **Show effects table with confidence intervals and p-values** is checked. This will provide not only the adjusted p-values for each comparison but also the confidence intervals for the mean difference, enriching the analysis. Click **OK** to execute [Tukey's method](#).

The image shows the 'pwmean - Pairwise comparisons of means' dialog box in Stata. It has three tabs: 'Main', 'if/in', and 'Reporting'. The 'Reporting' tab is selected. In the 'Confidence level' dropdown, '95' is selected. There are several checkboxes: 'Suppress effects tables' (unselected), 'Effects tables' (selected), 'Show effects table with confidence intervals' (unselected), 'Show effects table with p-values' (unselected), 'Show effects table with confidence intervals and p-values' (checked), 'Report means and confidence intervals' (unselected), 'Report means and group codes' (unselected), and 'Sort the means/differences in each term' (unselected). A 'Set table formats' button is located below the checkboxes. At the bottom, there are buttons for '?', a circular arrow, a document icon, 'OK', 'Cancel', and 'Submit'.

pwmean - Pairwise comparisons of means

Main if/in Reporting

95 Confidence level

☐ Suppress effects tables

☒ Effects tables

☐ Show effects table with confidence intervals

☐ Show effects table with p-values

☒ Show effects table with confidence intervals and p-values

☐ Report means and confidence intervals

☐ Report means and group codes

☐ Sort the means/differences in each term

Set table formats

? ↻ 📄 OK Cancel Submit

The resulting table provides a systematic comparison of every possible pair of drug groups. Each row in the output represents the mean difference between two groups, along with the standard error, the t-statistic, and the adjusted **p-value** derived using [Tukey's method](#).

```
. pwmean systolic, over(drug) mcompare(tukey) effects
```

Pairwise comparisons of means with equal variances

```
over          : drug
```

	Number of Comparisons
drug	6

systolic	Contrast	Std. Err.	Tukey t	P> t	Tukey [95% Conf. Interval]
drug					
2 vs 1	-.5333333	3.914808	-0.14	0.999	-10.911 9.844337
3 vs 1	-17.31667	4.152281	-4.17	0.001	-28.32385 -6.309485
4 vs 1	-12.56667	3.853153	-3.26	0.010	-22.7809 -2.352435
3 vs 2	-16.78333	4.152281	-4.04	0.001	-27.79051 -5.776152
4 vs 2	-12.03333	3.853153	-3.12	0.015	-22.24757 -1.819101
4 vs 3	4.75	4.094204	1.16	0.654	-6.103227 15.60323

Interpreting the **p-values** from this post-hoc table reveals which specific pairs demonstrated significant differences in the mean change of systolic blood pressure. For instance, the comparison between Drug 2 and Drug 1 yields an adjusted p-value of **0.999**. Since this value is far greater than 0.05, we conclude there is no statistically significant difference between these two specific drug groups. Conversely, we observe several pairwise comparisons where the p-value falls below the 0.05 threshold, indicating significant differences:

Drug 3 versus Drug 1: p-value = **0.001**

Drug 4 versus Drug 1: p-value = **0.010**

Drug 3 versus Drug 2: p-value = **0.001**

Drug 4 versus Drug 2: p-value = **0.015**

These results confirm that Drugs 3 and 4 were significantly more effective in changing systolic blood pressure compared to Drugs 1 and 2, while the differences between Drug 1 and 2, and between Drug 3 and 4, were non-significant.

Reporting Comprehensive ANOVA Findings

The final stage of the analysis involves summarizing the statistical findings in a clear, standardized format, typically adhering to guidelines such as those provided by the American Psychological

Association (APA). A complete report should include descriptive statistics for each group, the results of the omnibus ANOVA test, and the specific findings from the post-hoc multiple comparisons.

Step 6: Report the results.

The initial component of the report must detail the descriptive statistics, which include the sample size (n), mean, and standard deviation (SD) for the response variable within each level of the factor. This provides context for the subsequent inferential statistics. The table below, derived from the Stata output, summarizes these essential metrics:

Drug Used	Summary of Increment in Systolic B.P.		
	Mean	Std. Dev.	Freq.
1	26.066667	11.677002	15
2	25.533333	11.61813	15
3	8.75	10.0193	12
4	13.5	9.3238047	16
Total	18.87931	12.800874	58

Following the descriptive summary, the results of the [ANOVA](#) F-test are reported. This statement confirms whether the overall factor (drug type) had a significant effect. The report format includes the F-statistic, the degrees of freedom (Between groups, Within groups), and the exact **p-value**. A [One-Way ANOVA](#) was performed to determine if the four different types of drugs had differential impacts on systolic blood pressure. The analysis revealed a statistically significant difference between at least two of the drug groups, **F(3, 54) = 9.09, p = 0.001**.

Finally, the findings from the post-hoc analysis must be presented to delineate the specific nature of the differences. Based on [Tukey's method](#) for multiple comparisons, the change in systolic blood pressure was statistically significantly higher for Drug 3 compared to Drug 1 (Mean Difference = 17.32, Standard Error = 4.15, p = 0.001), and significantly higher for Drug 3 compared to Drug 2 (Mean Difference = 16.78, SE = 4.15, p = 0.001). Significant differences were also found between Drug 4 and Drug 1 (Mean Difference = 12.57, SE = 3.85, p = 0.010), and between Drug 4 and Drug 2 (Mean Difference = 12.03, SE = 3.85, p = 0.015). Importantly, there was no statistically significant difference in efficacy detected between Drug 1 and Drug 2 (Mean Difference = 0.533, SE = 3.91, p = 0.999) or between Drug 3 and Drug 4 (Mean Difference = 4.75, SE = 4.09, p = 0.654).